Language ideology in the contemporary Italian speech community:

A semantic vector space approach to the study of language attitudes in Italy

Stefano De Pascale, Dirk Speelman, Stefania Marzo

RU Quantitative Lexicology and Variational Linguistics

Overview

1. Language ideology and Cognitive Linguistics
    1. Cognitive Sociolinguistics
    2. Cognitive Contact Linguistics
2. Case study: Italian regional varieties
    1. Convergence between Italian and dialects
    2. Semantic Vector Space Models
    3. Semantic fields and language attitudes
3. Conclusions

1. Language ideology and Cognitive Linguistics

Social Psychology of Language (e.g. Giles)

- Little methodological innovation (matched guise experiments)
- Incompatible application of socialpsychological designs on language attitude research (Soukup 2013)

Critical Discourse Analysis (e.g. Fairclough)

- Lacking strong theoretical underpinning
- Methodological weakness (linguist's preconceptions) (Heylen, Wielfaert & Speelman 2013)

COGNITIVE SOCIOLINGUISTICS

1.1. Cognitive Sociolinguistics

- Cognitive Sociolinguistics as appropriate theoretical framework

Usage-based perspective

Centrality of meaning

Quantitative methods

Sociocultural language variation

Applied areas of linguistic investigation (Dirven, Polzenhagen & Wolf 2007)

1.2. Cognitive Contact Linguistics

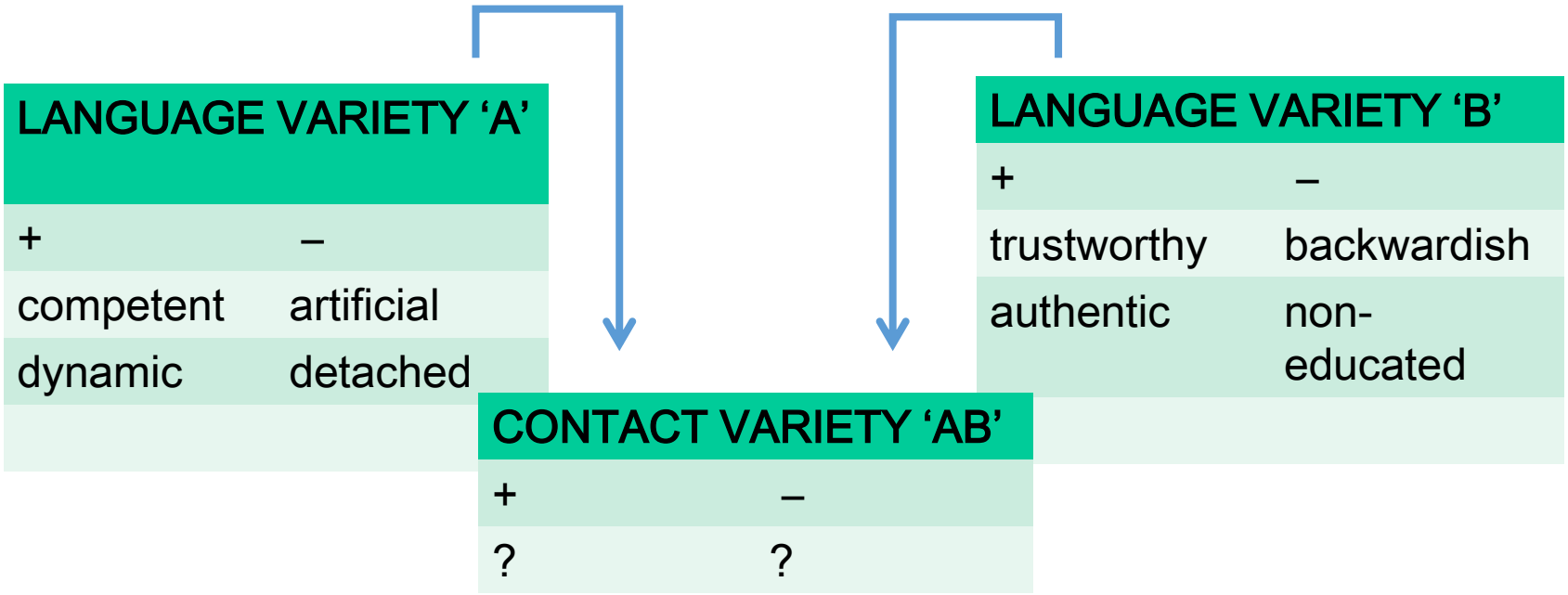
- Cognitive Contact Linguistics and language ideology

- Tap into people's conceptualization of language variation
- Sociocognitive correlates of linguistic outcomes of contact-induced change

- Does the hybrid nature of contact varieties reflect a hybrid language attitude architecture?
- How can we assess the (differential) contribution of the languages in contact/converging languages)?

1.2. Cognitive Contact Linguistics

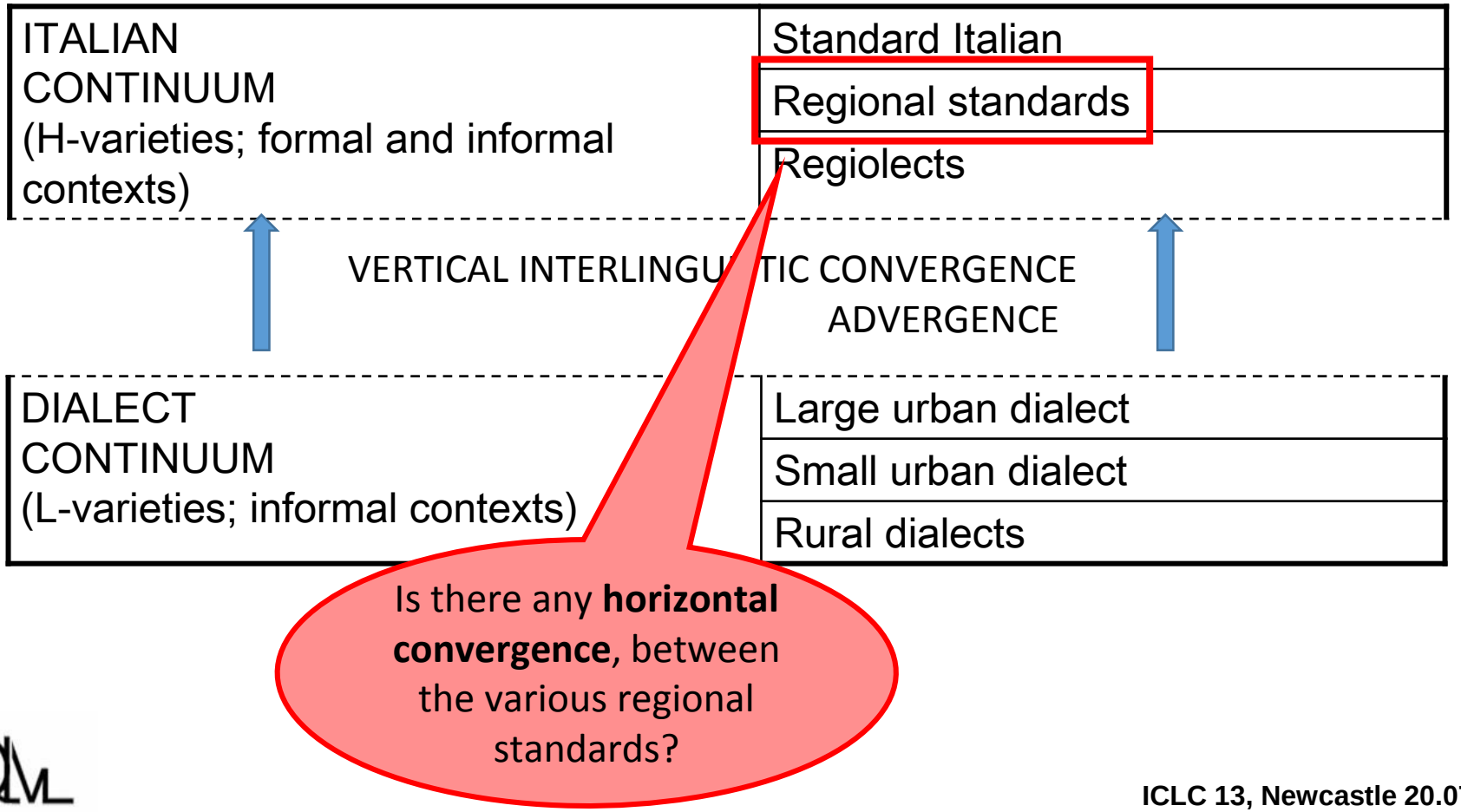
- Interesting potential interplay of rather complementary ideologies and stereotypes



2 Case study: Italian regional varieties

- Standard Italian:
  - Amended literary Florentine of the 14th century
  - Until the 20th century mainly a written language, LEARNED by most elites
- Italo-Romance dialects:
  - 5 systems scattered across Italy, very often mutually unintelligible
  - The language of everyday communication
- Turning point: socioeconomical changes in the 50's and 60's
  - Success of mass media, increased mobility, improved education
  - Standard Italian gains access to domains formely reserved to the dialects

2.1. Convergence between Italian and dialects(Cerruti & Regis 2014)



2.1. Convergence between Italian and dialects

- Evidence has shown that there is (Poletto 2009, Cerruti 2011)
- Younger speakers tend to adopt regional features from regional varieties other than their own

- The horizontality of this convergence is misleading, because it involves a social, and hence a "vertical" dimension as well.
  - Regional standards differ widely in prestige (Baroni 1983, Galli de' Paratesi 1984)
  - The more a variety is perceived to be closer to Standard Italian, the higher its prestige will be

2.1. Convergence between Italian and dialects

- Interesting potential interplay of rather complementary ideologies and stereotypes

STANDARD ITALIAN	
+	–
competent	artificial
dynamic	detached

REGIONAL ITALIAN VARIETY	
+	–
?	?

LOCAL/REGIONAL DIALECT	
+	–
trustworthy	backwardish
authentic	non-educated

2.2. Semantic Vector Space Models

Corpus-based analysis of experimentally elicited keywords

- Free Response Experiment
- 207 participants, mostly from the region Campania
- "Give the first 3 adjectives that come to mind for the following varieties:
  - Milanese Italian
  - Florentine Italian
  - Roman Italian
  - Neapolitan Italian"

2.2. Semantic Vector Space Models

- Distributional Hypothesis (Harris 1954)
- “You shall know a word by the company it keeps” (Firth 1957)

- Semantic similarity = central concept in distributional semantics
- Words that share the same linguistic context have similar meanings
- Large-scale collocation analysis (corpus-based!)

2.2. Semantic Vector Space Models

- STEP 1: Creation of a *term-by-document matrix*:
  - ROWS = *Terms*: keyword types (e.g. melodious, peasanty, modern)
  - COLUMNS = *Documents:* webpages from the ItWac corpus (Baroni et al. 2009)
  - CELLS/VECTORS = occurrence (1) or non-occurrence (0) of a keyword in a webpage
- STEP 2: Creation of a *item-by-item matrix* (= dissimilarity matrix)
  - Similarity measure: corrected Jaccard-index
- STEP 3: Cluster analysis

2.2. Semantic Vector Space Models

- STEP 1: Creation of a *term-by-document matrix*:
  - ROWS = *Terms*: keyword types (e.g. melodious, peasanty, modern)
  - COLUMNS = *Documents:* webpages from the ItWac corpus (Baroni et al. 2009)
  - CELLS/VECTORS = occurrence (1) or non-occurrence (0) of a keyword in a webpage
- STEP 2: Creation of a *item-by-item matrix* (= dissimilarity matrix)
  - Similarity measure: corrected Jaccard-index
- STEP 3: Cluster analysis

2.2. Semantic Vector Space Models

	Wp1	Wp2	Wp3	Wp4	Wp5	Wp6	Wp7
melodious	1	0	1	1	0	0	0
peasanty	1	0	0	0	1	1	1
modern	0	1	0	0	1	1	1

2.2. Semantic Vector Space Models

- STEP 1: Creation of a *term-by-document matrix*:
  - ROWS = *Terms*: 544 keyword types (e.g. melodious, peasanty, modern)
  - COLUMNS = *Documents:* webpages from the ItWac corpus (Baroni et al. 2009)
  - CELLS/VECTORS = occurrence (1) or non-occurrence (0) of a keyword in a webpage
- STEP 2: Creation of a *item-by-item matrix* (= dissimilarity matrix)
  - Similarity measure: corrected Jaccard-index
- STEP 3: Cluster analysis

2.2. Semantic Vector Space Models

- Corrected dissimilarity coefficients
  - Based on the rank order of the original coefficients
  - Double log-transformations of those ranks

	melodious	peasanty	modern
melodious	0	0,11	0,17
peasanty	0,11	0	0,11
modern	0,17	0,11	0

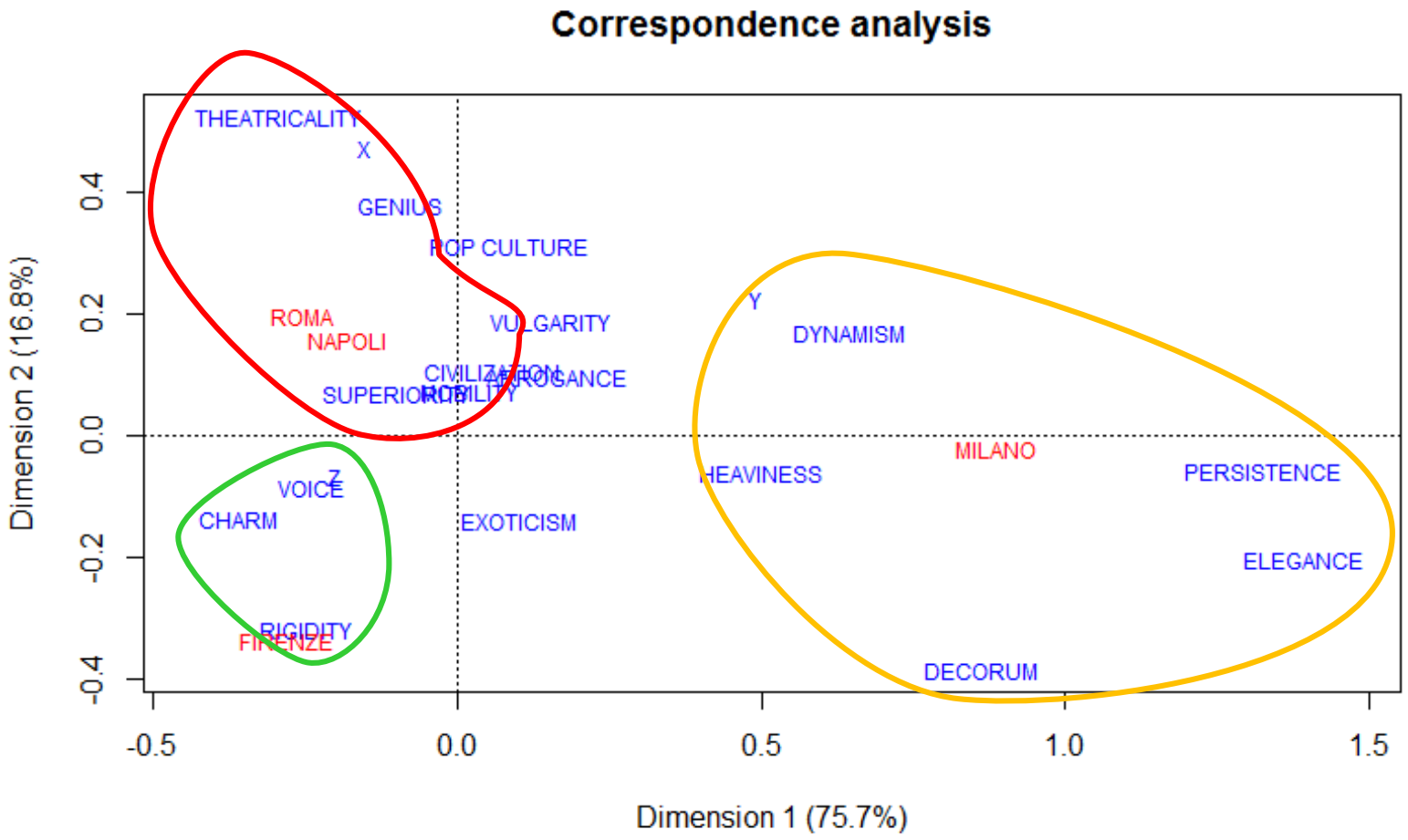
2.2. Semantic Vector Space Models

- STEP 1: Creation of a *term-by-document matrix*:
  - ROWS = *Terms*: keyword types (e.g. melodious, peasanty, modern)
  - COLUMNS = *Documents:* webpages from the ItWac corpus (Baroni et al. 2009)
  - CELLS/VECTORS = occurrence (1) or non-occurrence (0) of a keyword in a webpage
- STEP 2: Creation of a *item-by-item matrix* (= dissimilarity matrix)
  - Similarity measure: corrected Jaccard-index
- STEP 3: Cluster analysis

2.3. Semantic fields and language attitudes

- K-medoids clustering
  - Identify 20 clusters of related keywords = semantic fields
  - ⚠ semantic similarity ≠ semantic relatedness (Peirsman 2008)
  - Able to identify only 17 of 20 clusters
  - Naming often based on the most central members of the cluster
  - Correspondence analysis to visualize the correlation of regional varieties and their associated semantic fields

2.3. Semantic fields and language attitudes



2.3. Semantic fields and language attitudes

- Plot distances between varieties seem to reflect perceived geographical distances between those varieties, from the viewpoint of southern participants

- Milanese Italian in isolated position
- Roman Italian closer to Neapolitan Italian than to Florentine Italian

- After centuries of strong linguistic bond with Florence, Rome seems to rediscover its ancient linguistic southern roots (Cortelazzo 1974)

2.3. Semantic fields and language attitudes

- Milanese Italian: stereotype of the “Homo economicus”
  - Dynamism: boring/pasty BUT ALSO talkative/fluent
- Florentine Italian: admiration, respect
  - Almost all positive adjectives
  - Association with rhetorical and pronunciation qualities
- Roman Italian: superiority and civilization
- Neapolitan Italian: stereotype of the “Romantic Hero”
  - Unique, melodramatic, exaggerated
  - Also popular culture (often negative adjectives)

3. Conclusions

DESCRIPTIVE:

- Conceptualization of language varieties tend to follow a geographical pattern of north-south division
- Varieties, perceived as linguistically closer, are described with similar semantic fields

- Milanese variety described by means of domains traditionally associated with the social status of speakers
- Neapolitan variety described by means of domains traditionally associated with the personality traits of speakers

3. Conclusions

METHODOLOGICAL:

- Synthesis of corpus-based and experiment-based approach
- Semantic analysis of language attitudes reveals a richer architecture than simple negative-positive evaluation

- Semantic Vector Space Models are parameter-rich
- Still a lot of work to do in order to refine these techniques and provide evaluations of different modulations

3. Conclusions

THEORETICAL:

- Language attitudes as driving force for language (de)standardization
- *Standard Language Ideology in Contemporary Europe* (SLICE, University of Copenhagen)

- Assessing the direction of contact-induced language change by means of attitudinal and linguistic data
- Follow-up research project: corpus-driven, lectometrical study of Italian standardization dynamics

for further information:

stedepascale@gmail.com

dirk.speelman@arts.kuleuven.be

stefania.marzo@arts.kuleuven.be

http://wwwling.arts.kuleuven.be/qlvl

Bibliography

- Baroni, Marco, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. "The WaCky Wide Web: A collection of very large linguistically processed Web-crawled corpora." *Journal of Language Resources and Evaluation* 43 (3):209-226.
- Baroni, Maria R. 1983. Il linguaggio trasparente. Indagini psicolinguistica su chi parla e chi ascolta. Bologna: Il Mulino.
- Cerruti, Massimo. 2011. "Regional varieties of Italian in the linguistic repertoire." *International Journal of the Sociology of Language* 210:9-28.
- Cerruti, Massimo, and Riccardo Regis. 2014. "Standardization patterns and dialect/standard convergence: A northwestern Italian perspective." *Language in Society*43:83-111.
- Cortelazzo, Manlio. 1974. "Prospettive di studio dell'italiano regionale." In *Italiano d'oggi. Lingua non letteraria e lingue speciali*, edited by Mario Wandruszka, Manlio Cortelazzo and Maurizio Dardano, 19-33. Torino: LINT.
- Dirven, René, Frank Polzenhagen, and Hans-Georg Wolf. 2007. "Cognitive Linguistics, Ideology and Critical Discourse Analysis." In *The Oxford Handbook of Cognitive Linguistics*, edited by Dirk Geeraerts and Hubert Cuyckens, 1222-1241. New York: Oxford University Press.

Bibliography

- Firth, John. 1957. *Paper in Linguistics 1934-1951*. Oxford: Oxford University Press.
- Galli de' Paratesi, Nora. 1984. Lingua toscana in bocca ambrosiana. Tendenze verso l'italiano standard: un'inchiesta sociolinguistica. Bologna: Il Mulino.
- Harris, Zellig. 1954. "Distributional Structure." *Word* 10 (2/3):146-162.
- Heylen, Kris, Thomas Wielfaert, and Dirk Speelman. 2013. "Tracking Immigration Discourse through Time: A Semantic Vector Space Approach to Discourse Analysis." International Cognitive Linguistics Conference (ICLC 12), Edmonton, Alberta, Canada, 23-28 June 2015.
- Peirsman, Yves. 2008. "Word Space Models of Semantic Similarity and Relatedness." Proceedings of the ESSLLI-2008 Student Session, Hamburg, Germany.
- Poletto, Cecilia. 2009. "I costrutti verbo+preposizione: l'interferenza tra veneto e italiano regionale." In *Italiano, italiani regionali e dialetti*, edited by Anna Cardinaletti and Nicola Munaro, 155-172. Milano: Franco Angeli.
- Soukup, Barbara. 2013. "The measurement of language attitudes - a reappraisal from a constructionist perspective." In *Language (de)standardisation in Late Modern Europe: Experimental Studies*, edited by Tore Kristiansen and Stefano Grondelaers, 251-266. Oslo: Novus Press.